REPORT DOCUMENTATION PAGE		*Form Approved*
		OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
24-09-2015	Final	26 June 2014 – 25 June 2015

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
(134144) Robust Multimodal Cognitive Load Measurement II	FA2386-14-1-0007
	5b. GRANT NUMBER
	Grant 14IOA110 AOARD-134144
	5c. PROGRAM ELEMENT NUMBER
	61102F

6. AUTHOR(S)	5d. PROJECT NUMBER
Dr. Fang Chen	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
National ICT Australia (NICTA) Level 5, 13 Garden Street Eveleigh, NSW Sydney 2015 Australia	N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
AOARD UNIT 45002 APO AP 96338-5002	AFRL/AFOSR/IOA(AOARD)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
	AOARD-134144

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution A: Approved for public release. Distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report summarizes the important research activities, study results and research accomplishments out of the Robust Multimodal Cognitive Load Measurement II (RMCLM) project in the past one-year period. The objective of this project includes research of the fundamental issues related to the use of multiple input modalities and their fusion to enable robust and automatic cognitive load measurement (CLM) in the real world. Firstly, we examined the multimodal behavioral model to include mouse interactivity streams and a modified sliding window implementation for CLM. The contemplation-style and hesitation-style features are used in the study. The effects of cognitive load on trust are also investigated under two different conditions: low load and high load by analyzing behavioral signals such as mouse movement and physiological signals such as Galvanic skin response (GSR). Altogether, in the part one-year period, we had carried out CLM study with behavioral modalities such as mouse interactivity streams. The relations between cognitive load and trust were also analyzed with the use of mouse activities and GSR values.

15. SUBJECT TERMS
Behavioural Science, Cognitive Psychology

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Brian Lutz, Lt Col, USAF
U	U	U	SAR	8	**19b. TELEPHONE NUMBER** *(Include area code)* +81-80-5932-3242

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Final Report for AOARD Grant FA2386-14-1-0007 AOARD134144

Robust Multimodal Cognitive Load Measurement II

June 24, 2015

Name of Principal Investigators (PI and Co-PIs): Fang Chen
- e-mail address : fang.chen@nicta.com.au
- Institution : National ICT Australia (NICTA)
- Mailing Address : Level 5, 13 Garden Street, Eveleigh NSW 2015, Australia
- Phone : +61 2 9376 2101
- Fax : +61 2 9376 2025

Period of Performance: June/26/2014 – June/24/2015

Abstract: This report summarizes the important research activities, study results and research accomplishments out of the Robust Multimodal Cognitive Load Measurement II (RMCLM) project in the past one-year period. The objective of this project includes research of the fundamental issues related to the use of multiple input modalities and their fusion to enable robust and automatic cognitive load measurement (CLM) in the real world. Firstly, we examined the multimodal behavioral model to include mouse interactivity streams and a modified sliding window implementation for CLM. The contemplation-style and hesitation-style features are used in the study. The effects of cognitive load on trust are also investigated under two different conditions: low load and high load by analysing behavioural signals such as mouse movement and physiological signals such as Galvanic skin response (GSR). Altogether, in the part one-year period, we had carried out CLM study with behavioral modalities such as mouse interactivity streams. The relations between cognitive load and trust were also analyzed with the use of mouse activities and GSR values.

List of Publications
a) Papers published in peer-reviewed journals

1. J. Zhou, J. Sun, F. Chen, Y. Wang, R. Taib, A. Khawaji, and Z. Li, "Measurable Decision Making with GSR and Pupillary Analysis for Intelligent User Interfaces", ACM Transactions on Computer-Human Interaction (ToCHI), vol. 21, no. 6, article no. 33, 2015.

b) Papers published in peer-reviewed conference proceedings:

2. J. Zhou, C. Bridon, F. Chen, A. Khawaji, and Y. Wang, "Be Informed and Be Involved: Effects of Uncertainty and Correlation on User's Confidence in Decision Making", SIGCHI Conference on Human Factors in Computing Systems (CHI2015), 2015.
3. A. Khawaji, J. Zhou, F. Chen, and N. Marcus, "Using Galvanic Skin Response (GSR) to Measure Trust and Cognitive Load in the Text-Chat Environment", SIGCHI Conference on Human Factors in Computing Systems (CHI2015), 2015.
4. S. Arshad, Y. Wang, and F. Chen, "Interactive Mouse Stream as Real-Time Indicator of User's Cognitive Load", SIGCHI Conference on Human Factors in Computing Systems (CHI2015), 2015.
5. A. Khawaji, F. Chen, J. Zhou and N. Marcus, "Trust and Cognitive Load in the Text-Chat Environment: The Role of Mouse Movement", Proceedings of OzCHI 2014.

Contents

Attachment A	Measurable Decision Making with GSR and Pupillary Analysis for Intelligent User Interface
Attachment B	Be Informed and Be Involved: Effects of Uncertainty and Correlation on User's Confidence in Decision Making
Attachment C	Using Galvanic Skin Response (GSR) to Measure Trust and Cognitive Load in the Text-Chat Environment
Attachment D	Interactive Mouse Stream as Real-Time Indicator of User's Cognitive Load
Attachment E	Trust and Cognitive Load in the Text-Chat Environment: The Role of Mouse Movement

1. Introduction

Cognitive load is the load on working memory experienced by user when undertaking a learning or mission critical task. Managing cognitive load is desirable to enhance productivity (in learning settings) and avoiding fatal scenarios in time critical problems. However, measuring cognitive load remains the holy grail of cognitive load theory. A number of approaches exist for measuring cognitive load but with associated limitations. Physiological and behavioral approaches appear more suited for real time scenarios. Both assume that higher mental effort will cause observable involuntary/voluntary changes in human physiological/behavioral measures. Furthermore, various factors may affect cognitive load, such as trust, which can be revealed through physiological and behavioral signals.

This project has focused on behavioral measures especially mouse movement in indexing cognitive load. The affecting factors on cognitive load such as trust were also investigated. We also studied how physiological signals such as Galvanic Skin Response (GSR) varied because of affecting factors.

2. Mouse Activity for CLM

Mouse dynamics is an actively explored behavioral biometric research area for both static and continuous user authentication. We were inspired by trajectory analysis techniques being investigated for user's continuous authentication. Mouse curve (i.e. mouse movements with little or no pause between them) features (like length, curvature and inflection) were used to classify and develop classification histograms that would be representative of an individual's typical mouse usage. This problem environment bears strong resemblance to that of continuous cognitive load monitoring scenario. However, the key problem, for biometric research, remains to minimize and handle intelligently the effects of within-user behavior variability (for extended periods of mouse usage). Luckily, this presents an opportunity for our research, as we already presume change in user behavior under varying cognitive load conditions and behavioral biometric findings support our case.

Experiment: We use data from a larger multitier experiment that was designed to study effects of cognitive load on organizational trust. Cognitive load was varied using standard dual-task design. Low load condition comprised of a sole primary task, whereas high load condition included an additional secondary task that periodically popped into user's view and conditionally required a classification action. The experiment was conducted through a simulated computer-based platform to screen applicants for a human resource department. Every subject interacted with this tool; studied the information provided and completed the AIB (Ability, Integrity & Benevolence) trust indicator related task while user interaction data (including mouse activity) was recorded. All subjects completed both (low and high cognitive) load conditions in a repeated measures design. The scope of current analysis is limited to processing user mouse activity streams captured from 27 subjects. Each subject data stream was approximately balanced (in terms of data points) for both load level mouse activity.

Mouse Activity Features: A pause in mouse activity refers to the interval between two consecutive mouse events. It has been argued that contemplation-style and hesitation-style interval categories hold significant promise for detecting high cognitive load on working memory. Contemplation-style interval corresponds to the more clearly observable break in user input activity that ranges from about 1 second to 5 seconds. This type of interval shows change patterns similar to those previously observed for speech pause features. On the other hand, hesitation-style intervals typically range from more than one-tenth of a second to one

second apart. These correspond to subtle variations in user input behavior, that may be interpreted as 'hesitant' or 'cautious' due to high cognitive load on working memory.

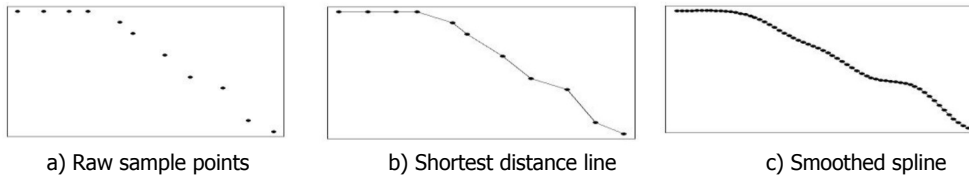| a) Raw sample points | b) Shortest distance line | c) Smoothed spline |

Figure 1. Estimating user behavior from raw mouse movement sample points.

The problem of cognitive load detection from multimodal data streams was formulated as the problem of detecting concept drift from data streams. Sliding windows technique solves, usefully, the problem of detecting user's cognitive load in live scenarios. Figure 2 shows the block diagram of sliding windows implementation.
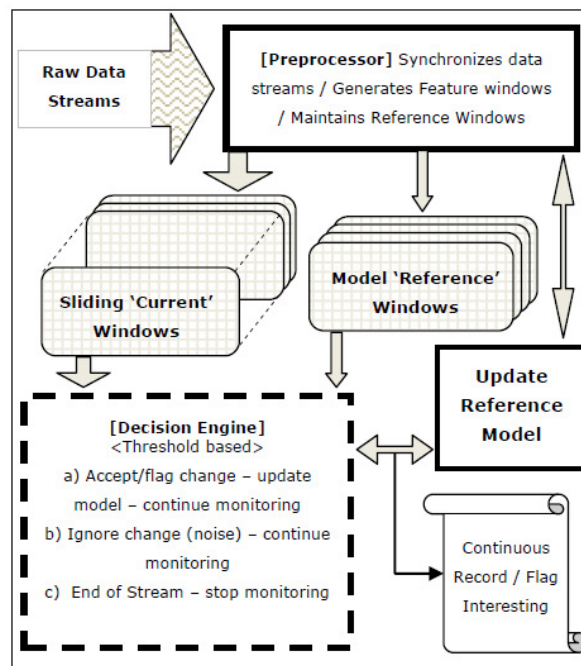

Figure 2. Block diagram of sliding windows implementation.

Results and Discussion: Experimental results show that positive predicted value (PPV) results of load variations detected for the 27 subject streams. Every stream had grounded 19 true shifts and a mean value of 8.2 true positive detections was achieved (which is reasonably good considering the fact that reference model was updated dynamically after each detection, whereas earlier approaches had limited or no updating capability of learning from new data).

We find that learning from mouse interactivity and using sliding window further strengthens the multimodal framework for real time cognitive load detection.

3. Relations between Trust and Cognitive Load

Trust refers to a situation when someone can predict how others will behave and what will occur from their behaviors. It is also defined as "*a willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party*".

4

Cognitive load refers to the amount of mental load imposed on a human's working memory when a person attempts to accomplish a task. An increased amount of new information has a significant impact on the way people behave. Previous research showed that if people were given extra time (15 minutes) to chat in the text chat environment, it built a higher level of trust between them when they chatted again later via the same medium compared with people who didn't chat for additional time. This finding raises a question about what happens to the trust between people in this medium when their attention is distracted from the communication. To explore this question, we examine the effects of cognitive load on trust. under two different conditions: low load and high load tasks.

3.1. The Role of Mouse Movement
Experiment:
The data were collected using the DayTrader task which requires players to communicate with each other to play an investment game. This investment game follows the rules of the Prisoner's Dilemma game. To obtain high and satisfying rewards, players must trust each other. In this study, each participant chatted and played with one other participant. Each participant was exposed to two cognitive load conditions, low load and high load, but only during their chats with each other. The cognitive load was introduced by asking participants to sum random numbers in their heads, without using pen and paper or a calculator, and enter the total of the numbers at the end of each session.

Twenty participants were recruited for this study (13 males and 7 females, aged between 22 and 40). All the participants were university students and none of them had met each other prior to the task. The participants were randomly assigned to chat with their partner.

In this study, the movements of the mouse cursor in the graphical user interface were recorded for analysis.

Mouse Movement Measures: During the chat sessions, when the mouse cursor moved, the time stamps and coordinates (X, Y) of the mouse cursor were recorded. For each two sequential pairs of coordinates (X, Y) and (X, Y) which constitute a line (in other words, a movement), we called these two A (AX, AY) and B (BX, BY) to carry out the calculation. We calculated a set of measures for the mouse movements. These measures are: distance, slope, movement count, and duration. Table 1 shows the summary of mouse movement measures under two cognitive load conditions.

Table 1. Summary of mouse movement measures when cognitive load is low and high.

Measure	High Load Mean(SD)	Low Load Mean(SD)	t	p-value
Distance (pixels)	16954(6549)	25026(10055)	3.93	0.004
Positive Slope (+)	478(299)	861(497)	3.74	0.005
Negative Slope (-)	1058(440)	1490(450)	4.32	0.003
Movement Count (movements)	556(236)	809(378)	2.89	0.015
Duration (seconds)	655(231)	710(172)	0.90	0.401

Results and Discussion: The experimental results show that the distance travelled by the participants' mouse was significantly higher (p=0.004) when the participants' mental load was low (M=25026 pixels) compared with when the participants' mental load was high (M=16954 pixels). Similarly, the total steepness of lines for positive and negative slopes increased significantly (p=0.005, p=0.003), from when the participants' mental load was high (M=+478, M=-1058) to when the participants' mental load was low (M=+861, M=-1490), respectively. In addition, the total number of movements significantly increased (p=0.015), from when the participants' mental load was high (M=556 movements) to when the participants' mental load was low (M=809 movements). Finally, the total length of time that the mouse cursor stopped failed to show significant results (p=0.401). It was noted that when the participants summed

large random numbers, they faced an extreme load on their working memories which was reflected directly in their attitudes and feelings toward their partners and their way of moving the mouse.

3.2. GSR as a Measure to Show Relations between Trust and Cognitive Load

Experiment: The participants were invited to do an investment game in pairs with the DayTrader task, where the partners were allowed to chat with each other using instant messaging and play the investment game in the same window. The rules of this investment game were taken from the Prisoner's Dilemma task where the payoff resulting from the investment game was used to measure trust (when an increase in the group payoff indicates an increase in trust and vice versa).

Twenty-eight students from NSW University and NICTA organization were recruited for this study, their ages ranging from 18 to 40 years (18 males and 10 females). Each participant was assigned randomly to one partner during the experiment, and in total, there were fourteen pairs. The experiment was designed as follows: 2 trust levels (low/high) x 2 cognitive load levels (low/high) in a mixed design (four conditions): Low Trust-High Cognitive Load (LTHCL), Low Trust-Low Cognitive Load (LTLCL), High Trust-High Cognitive Load (HTHCL) and High Trust-Low Cognitive Load (HTLCL).

The GSR signals were collected during the DayTrader task.

Results and Discussion: The experimental results show the subjects' trust, as measured by the group payoff, was highest when they met each other face-to-face before they started chatting and after having read a paragraph encouraging trust in the instructions of the task and also the subjects' cognitive load was highest when they summed large random numbers. In this study, we examined two GSR features: the average of GSR values and the average of peaks of GSR values. The results show that the average of the GSR values is decreased in the HTLCL condition more than the other conditions (see Figure 3). A two-tailed two-way ANOVA was conducted to examine the interaction effect between trust and cognitive load on the averages of GSR and the results showed significant interaction between them ($F=5.3$, $p=0.02$). The results of the averages of peaks are as follows, from largest to smallest: LTLCL ($M=2.52E-5$, $SD=9.83E-6$), LTHCL ($M=2.3E-5$, $SD=8.5E-6$), HTHCL ($M=2.2E-5$, $SD=1.3E-5$) and HTLCL ($M=1.4E-5$, $SD=5.2E-6$) (see Figure 4). A two-tailed two-way ANOVA was conducted to examine the interaction effect between trust and cognitive load on the averages of peaks and the results showed significant interaction between them ($F=5.4$, $p=0.02$).
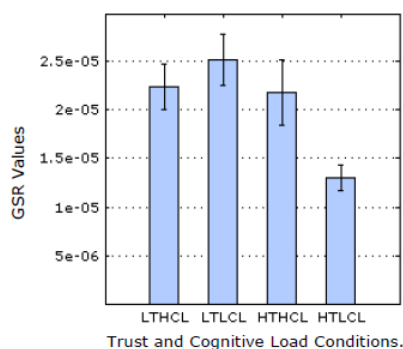

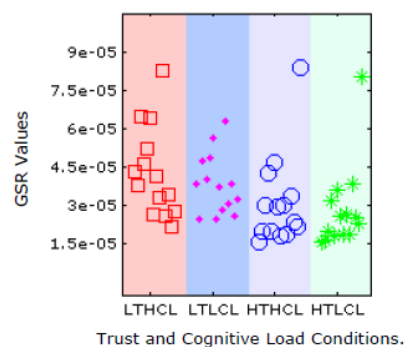Figure 3. The average of GSR values.

Figure 4. The average of peak GSR values.

The results show that in a low cognitive load situation, that is in LTLCL and HTLCL conditions, GSR can be used to measure the level of interpersonal trust, while in a high trust situation, that is in HTLCL and HTHCL conditions, GSR can be used to measure the level of cognitive load.

4. Conclusions and Future Work

This research carried out CLM study of behavioral measures especially mouse activities, and the relations between trust and cognitive load by analyzing mouse movement and GSR signals.

In the mouse activity based CLM, we showed that good progress has been made in direction of using multimodal behavior and interactivity as indicator of cognitive load. We extended these efforts by enhancing the multimodal behavioral model to include mouse interactivity streams and a modified sliding window implementation. As we continue to experiment with several variations on modified adaptive sliding windows, the technical feasibility of learning from streams is helping us understand the limitations for measuring real time cognitive load indices.

The mouse activity was also investigated to find relations between trust and cognitive load. The results showed that when the participants summed large random numbers, they faced an extreme load on their working memories which was reflected directly in their attitudes and feelings toward their partners and their way of moving the mouse. The trust results revealed support for our hypothesis that the level of cognitive load affects the building of trust when people communicate in the chat medium.

The relations between trust and cognitive load were also investigated with GSR signals. The results showed that GSR values were significantly lower when both trust is high and cognitive load is low. Moreover, this study has provided promising findings in relation to indicators for determining the level of trust and cognitive load. Specifically, the results of GSR values show that in a low cognitive load situation, that is in LTLCL and HTLCL conditions, GSR can be used to measure the level of interpersonal trust, while in a high trust situation, that is in HTLCL and HTHCL conditions, GSR can be used to measure the level of cognitive load. The GSR values were at their lowest level when the participants' trust is high and they were exposed to a low cognitive load.

Future work will include setting up the cognitive load dynamic feedback loop by adaptively changing cognitive load in human-machine systems to allow cognitive load level be an appropriate level (not too high, not too low) in order to improve interaction performance. We will also investigate how different modalities of CLM can be fused in order to improve the CLM accuracy. Other works include the development of new features for different modalities and more accurate machine learning models for CLM.